A PROBLEM IN THE SEQUENTIAL DESIGN OF EXPERIMENTS

Richard Bellman

P-586

20 October 1954

The RAND Corporation
1700 MAIN ST. · SANTA MONICA · CALIFORNIA

SUMMARY

~~We consider~~ The problem of determining an optimal testing policy where ~~we~~ one simultaneously gains and learns for the case where the outcome of one choice is known and the other is subject to a known à priori distribution.

Results of Johnson and Karlin, P-328, are obtained in a different way and extended. The methods used are applicable to more general processes.

A PROBLEM IN THE SEQUENTIAL DESIGN OF EXPERIMENTS

Richard Bellman

§1. Introduction

In two little-known papers written in 1933 and 1935, [14], [15], W. R. Thompson proposed the problem of determining on the basis of sequential analysis which of two drugs were superior.* The problem is a difficult one, and Thompson concentrated his efforts on the computation of the effects of a plausible policy, and on a Monte Carlo determination of the outcome.

A problem in the same general area was discussed by Mahalanobis, [11], [12], in connection with a sampling survey of the acreage under jute in Bengal.

An interesting exposition of the general problem is given by Robbins, [13], where further references may be found. The connection with the Wald theory of sequential analysis is discussed, and further problems in this field are presented.

The problem is also in the general field of "learning processes", where we must determine the structure of a process while carrying on an experiment, cf [8], [10], [13].

* We confess that we found these papers in the standard fashion, namely while thumbing through a journal containing another paper of interest.

In a recent paper, [9], Johnson and Karlin considered a particular version of the Thompson problem, essentially the case where one drug has known properties and the other unknown, and derived a number of interesting results concerning the structure of an optimal policy.

In this paper, we shall consider their problem and an analogous problem by means of a discussion of the functional equation derived from the process. Using techniques we have employed in various parts of the theory of dynamic programming, [1], [2], [4], [7], we shall determine the structure of the optimal policy and complete the Johnson-Karlin results in an essential detail.

In §2 we present a precise formulation of the problem we treat here. In §3 we derive the basic functional equation, with properties of the solution, existence, uniqueness and successive approximations discussed in §4. The next section contains a statement of the results we obtain concerning the structure of the optimal policy. In §6 we present a proof of these results. Finally in §7 we discuss the numerical computation of the solution based upon successive approximations.

The methods we employ are applicable to more general processes, as we shall show in further papers.

§2. Formulation of the Problem

Let us assume that we have two machines, unimaginatively called I and II, with the following properties. If machine I is used, there is a probability r of receiving a gain of one unit, and a probability 1-r of receiving nothing. If machine II is used, there is a corresponding probability of s.

Unfortunately these probabilities are not known. We do, however, possess an à priori probability distribution for their values, F(r,s).

We may now consider either a finite sequence of choices where we have n trials, or an unbounded process with a discount factor a for the value of a unit received one trial away. The infinite process is simpler analytically since it possesses an invariant aspect over time. Our methods are equally applicable to both types of processes.

The problem is now to determine the sequence of choices which maximizes the total expected return. This sequence is in general stochastic since the choice after any finite number of choices will depend upon the outcomes of the preceding choices.

In this paper, we shall consider only the simple case where r and s are uncorrelated, and even further we shall assume that s is known. Let F(r) be the distribution function for r in [0,1].

§3. The Basic Functional Equation

We shall utilize an analytic approach based upon a functional equation associated with the problem. Let us define

(1) $f_{m,n}(s)$ = expected return obtained using an optimal policy for an unbounded process after the first machine has had m successes and n failures.

Our fundamental assumption is the usual one that the new à priori distribution function after m successes and n failures on the first (unknown) machine is given by

$$(2) \quad dF_{mn}(r) = \frac{r^m(1-r)^n dF(r)}{\int_0^1 r^m(1-r)^n dF(r)}$$

On the basis of this assumption, an enumeration of outcomes yields the relation, if the first machine is chosen,

$$(3) \quad f_{m,n}(s) = \int_0^1 r dF_{mn}(r) \left[1 + af_{m+1,n}(s) \right] + \int_0^1 (1-r) dF_{mn}(r) \left[af_{m,n+1}(s) \right].$$

On the other hand, if the second (known) machine is chosen, we have

$$(4) \quad \begin{aligned} f_{m,n}(s) &= s + af_{m,n}(s) \\ &= s/(1-a). \end{aligned}$$

Hence we obtain the fundamental recurrence relation

$$(5) \quad f_{m,n}(s) = \text{Max} \left[\begin{array}{ll} \text{I:} & \int_0^1 r dF_{mn}(r) \left[1 + af_{m+1,n}(s) \right] \\ & + \int_0^1 (1-r) dF_{mn}(r) \left[af_{m,n+1}(s) \right], \\ \text{II:} & s/(1-a). \end{array} \right]$$

a typical functional equation in the theory of dynamic programming.

Let us now introduce some simplifying notation. Write

$$(6) \quad f_{m,n}(s) = f(m,n),$$

$$\int_0^1 r dF_{mn}(r) = b(m,n).$$

Then (5) takes the simpler form

$$(7) \quad f(m,n) = \text{Max} \begin{bmatrix} \text{I:} & b(m,n) \left[1 + af(m+1,n) \right] + a(1-b(m,n)f(m,n+1)) \\ \text{II:} & s/(1-a) \end{bmatrix}$$

for $m,n \geq 0$.

Let us note that $0 < a$, $s < 1$, and that $0 < b(m,n) < 1$ for $m,n \geq 0$.

§4. Existence and Uniqueness of Solution

Since our analysis of the structure of the optimal policy will be based upon a continued application of successive approximations to the system in (3.7), it is essential to have an existence and uniqueness theorem and information concerning the convergence of successive approximations.

The method we employ is equally applicable to other functional equations in dynamic programming and examples may be found in [1], [2], [4], [5], [7].

Theorem 1. *There is a unique solution to* (3.7), $\left\{f(m,n)\right\}$, *which is uniformly bounded by* $1/(1-a)$ *for all* m *and* $n \geq 0$. *This solution may be obtained as the limit as* $k \to \infty$ *of the sequence* $\left\{f_k(m,n)\right\}$, *defined recurrently by*

$$(1) \quad f_0(m,n) = g(m,m)$$

$$f_{k+1}(m,n) = T_{mn}(f_k), \; k = 0,1,2,\ldots, \; m,n \geq 0,$$

where we set

$$(2) \quad T_{mn}(f) = \text{Max} \begin{bmatrix} \text{I}: & b(m,n)(1 + af(m+1,n)) + (1-b(m,n)) \left[af(m,n+1) \right] \\ \text{II}: & s/(1-a) \end{bmatrix}$$

Here $\{g(m,n)\}$ may be any sequence uniformly bounded by $1/(1-a)$.

Proof: Let us define

$$(3) \quad u_I(f,m,n) = b(m,n) \left[1 + af(m+1,n) \right] + a \left[1-b(m,n) \right] f(m,n+1)$$

$$u_{II}(f,m,n) = s/(1-a)$$

Then for each $m,n \geq 0$, we have

$$(4) \quad f_{k+1}(m,n) = u_A(f_k,m,n),$$

where $A = \text{I}$ or II and the choice is dependent upon m,n and f_k.
Similarly

$$(5) \quad f_k(m,n) = u_B(f_{k-1},m,n),$$

where B may equal A.

In any case, we have, by virtue of the recurrence relation of (1), the inequalities

$$(6) \quad f_{k+1}(m,n) = u_A(f_k,m,n) \geq u_B(f_k,m,n)$$

$$f_k(m,n) = u_B(f_{k-1},m,n) \geq u_A(f_{k-1},m,n).$$

Hence

$$(7) \quad f_{k+1}(m,n) - f_k(m,n) \geq u_A(f_k,m,n) - u_A(f_{k-1},m,n)$$
$$\leq u_B(f_k,m,n) - u_B(f_{k-1},m,n)$$

These inequalities yield

$$(8) \quad |f_{k+1}(m,n) - f_k(m,n)| \leq \text{Max} \left[|u_I(f_k,m,n) - u_I(f_{k-1},m,n)|, \; |u_{II}(f_k,m,n) - u_{II}(f_{k-1},m,n)| \right],$$

or, using the analytic expression for u_I and u_{II},

$$(9) \quad |f_{k+1}(m,n) - f_k(m,n)| \leq ab(m,n)|f_k(m+1,n) - f_{k-1}(m+1,n)|$$
$$+ a(1-b(m,n))|f_k(m,n+1) - f_{k-1}(m,n+1)|$$
$$\leq a \, \text{Max} \left[|f_k(m+1,n) - f_{k-1}(m+1,n)|, \; |f_k(m,n+1) - f_{k-1}(m,n+1)| \right].$$

If we set

$$(10) \quad u_k = \operatorname*{Sup}_{m,n \geq 0} | f_k(m,n) - f_{k-1}(m,n) |,$$

the inequality in (9) yields

$$(11) \quad u_{k+1} \leq au_k.$$

From this it follows that the series

$$(12) \quad S(m,n) = \sum_{k=0}^{\infty} (f_{k+1}(m,n) - f_k(m,n)),$$

converges uniformly in m and n for $m,n \geq 0$, and that $f_k(m,n) \to f(m,n)$ as $k \to \infty$.

It is readily verified that $\{f_k(m,n)\}$ is uniformly bounded by $1/(1-a)$ for $0 \leq s \leq 1$ if this holds for $\{g(m,n)\}$.

To establish uniqueness, let $\{F(m,n)\}$ be another solution, uniformly bounded by $1/1-a$, or any fixed quantity.

Proceeding as in (4) - (9), we obtain the inequality

$$(13) \quad |f(m,n) - F(m,n)| \leq a \operatorname{Max} \left[|f(m+1,n) - F(m+1,n)|, \; |f(m,n+1) - F(m,n+1)| \right].$$

Setting

$$(14) \quad u = \sup_{m,n \geq 0} |f(m,n) - F(m,n)|,$$

the inequality in (13) yields

$$(15) \quad u \leq au,$$

and consequently the result that $u = 0$ or $f(m,n) = F(m,n)$.

§5. Statement of Results

Let us now state the results we shall prove concerning the structure of the solution of (3.7). Observe that it is the "policy", i.e. the value of s which dictates a choice of machine I or machine II which determines the solution.

Theorem 2. For each $m,n \geq 0$, there is a unique quantity $s(m,n)$ with the property that

$$(1) \quad (a) \quad f(m,n) = s/(1-a), \; 1 \geq s \geq s(m,n),$$

$$(b) \qquad = b(m,n)[\,1 + af(m+1,n)\,] + a(1-b(m,n))f(m,n+1),$$

$$0 \leq s \leq s(m,n).$$

The sequence $\{s(m,n)\}$ has the following properties

(2) $s(m+1,n) > s(m,n) > s(m,n+1)$,

and the sequence $\{f(m,n)\}$ similarly satisfies the relations

(3) $f(m+1,n) > f(m,n) > f(m,n+1)$

Analogous results hold for the case of a finite number of trials.

The proof which we shall present in the next section will be based on the method of successive approximations.

§6. Proof of Theorem 2

We shall approximate to the solution of the original equation by means of the sequence $\{f_k(m,n)\}$ defined as follows

(1) $f_0(m,n) = \text{Max}\ [\ b(m,n),\ s/(1-a)\ ]$,

$f_{k+1}(m,n) = T_{mn}(f_k)$, $k = 0,1,2,\ldots$, $m,n \geq 0$.

We wish to prove the following statements

(2) (a) $f_{k+1}(m,n) > f_k(m,n)$

(b) for all $k \geq 0$, there is a sequence $\{s_k(m,n)\}$ with the property that

(1) for $s \geq s_k(m,n)$, $f_{k+1}(m,n) = s/(1-a)$,

(2) for $s \leq s_k(m,n)$, $f_{k+1}(m,n) = u_I(f_k,m,n)$.

(c) $f_k(m+1,n) > f_k(m,n) > f_k(m,n+1)$,

(d) $s_k(m+1,n) > s_k(m,n) > s_k(m,n+1)$,

(e) $s_{k+1}(m,n) > s_k(m,n)$.

Let us begin with the case $k = 0$. Let $\{s_0(m,n)\}$ be the sequence determined by the equation

(3) $s/(1-a) = b(m,n)$.

Then (2b) is clearly true for $k = 0$. To obtain further relations we require the inequalities

(4) $b(m,n+1) < b(m,n) < b(m+1,n)$

for all $m,n \geq 0$.

The second inequality is equivalent to

$$(5) \quad \frac{\int_0^1 r^{m+1}(1-r)^n dF}{\int_0^1 r^m(1-r)^n dF} < \frac{\int_0^1 r^{m+2}(1-r)^n dF}{\int_0^1 r^{m+1}(1-r)^n dF}$$

or

$$(6) \quad (\int_0^1 r dG)^2 < (\int_0^1 dG)\,(\int_0^1 r^2 dG)$$

where $dG = r^m(1-r)^n dF$. This, however, is the Cauchy-Schwarz inequality

It may readily be verified that the first inequality is also equivalent to (6).

These inequalities, (4), yield (2c) and (2d) for $k = 0$.

Let us now begin the induction. We have

(7) $f_1(m,n) = T_{mn}(f_0)$.

Since $s_1(m,n)$ is determined by the equality of the two expressions in $T_{mn}(f_0)$, it is clear that

(8) $s_1(m,n) > s_0(m,n)$.

Let us now demonstrate the essential result that

(9) $s_1(m+1,n) > s_1(m,n) > s_1(m,n+1)$

Consider the equation for $s_1(m,n)$. We have, with $s = s_1(m,n)$,

(10) $\frac{s}{1-a} = b(m,n) + ab(m,n)f_0(m+1,n) + a(1-b(m,n))f_0(m,n+1)$.

However, since

(11) $s_1(m,n) > s_0(m,n) > s_0(m,n+1)$,

we have for this value of s

(12) $f_0(m,n+1) = s/(1-a)$

Hence (10) reduces to

(13) $\frac{s}{1-a} = b(m,n) + ab(m,n)f_0(m+1,n) + a(1-b(m,n))s/(1-a)$,

which yields

(14) $\frac{s}{1-a} = \frac{(1-a)b(m,n)(1 + af_0(m+1,n))}{1-a + ab(m,n)}$

Since $x/(1-a + ax)$ is monotone increasing x for $x \geq 0$, it follows that

$$(15)\quad \frac{(1-a)b(m+1,n)}{(1-a)+ab(m+1,n)} > \frac{(1-a)b(m,n)}{(1-a)+ab(m,n)}$$

Since $f_0(m+1,n)$ is monotone increasing in m for all n and s, it follows that the curve

$$(16)\quad u_{m+2}(s) = \frac{(1-a)b(m+1,n)}{(1-a)+ab(m+1,n)}\,(1+af_0(m+2,n)),$$

lies above the curve for $u_{m+1}(s)$. Hence $s_1(m+1,n) > s_1(m,n)$. The same proof shows that $s_1(m,n+1) < s_1(m,n)$.

The last step of the induction consists of the proof that

$$(17)\quad f_1(m+1,n) > f_1(m,n) > f_1(m,n+1).$$

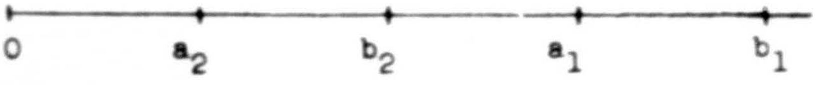
Consider the proof of the first of these inequalities. We have

$$(18)\quad f_1(m+1,n) = \text{Max}\left[\begin{array}{c} b(m+1,n) + ab(m+1,n)f_0(m+2,n) + a(1-b(m+1,n)) \\ \hfill f_0(m+1,n+1) \\ s/(1-a) \end{array}\right]$$

Let us set

$$(19)\quad \begin{array}{ll} a_1 = f_0(m+1,n), & a_2 = f_0(m,n+1) \\ b_1 = f_0(m+2,n), & b_2 = f_0(m+1,n+1) \\ \lambda = b(m,n), & \mu = b(m+1,n) \end{array}$$

The location of a_2,b_2,a_1,b_1 on the real axis is as follows:



by virtue of the inequalities holding for $f_0(m,n)$. Since $\mu > \lambda$, it is sufficient to show that the convex combination $\mu b_1 + (1-\mu)b_2$ is greater than or equal to the convex combination $\lambda a_1 + (1-\lambda)a_2$.

Consider the linear expression $E(\lambda) = \lambda a_1 + (1-\lambda)a_2$ for $0 \leq \lambda \leq \mu$. At $\lambda = \mu$, we have

(20) $E(\mu) = \mu a_1 + (1-\mu)a_2 < \mu b_1 + (1-\mu)b_2$

At $\lambda = 0$, we have

(21) $E(0) = a_2 < \mu b_1 + (1-\mu)b_2$

for any $\mu \geq 0$.

Consequently, for all values of λ in the interval $[0,\mu]$ we have

(22) $E(\lambda) < \mu b_1 + (1-\mu)b_2$.

Comparing the expression for $f_1(m+1,n)$ with that for $f_1(m,n)$ it follows that $f_1(m+1,n) > f_1(m,n)$. The inequality $f_1(m,n) > f_1(m,n+1)$ is derived similarly.

We now have all the details required for an inductive proof which proceeds from k and k+1 in precisely the fashion above.

Since the inequalities are valid for all k, they are valid for the limiting sequence $\left\{f(m,n)\right\}$, with strict inequality because of the strict inequality in the relation $b(m+1,n) > b(m,n)$.

§7. Discussion

It seems to be a very difficult problem to determine the precise analytic form of s_{mn}. Consequently the most efficient method of determining this sequence is probably by means of successive approximations starting with a suitable $f_0(m,n)$.

It is worth pointing out that in place of starting out with an initial approximation, $\{f_0(m,n)\}$, it is better to guess an initial policy, $\{s_{mn}\}$. It is simpler, and more natural, to choose a sequence of values rather than a sequence of functions. Furthermore, we have a much stronger feel for an approximate policy than we do for an approximate function, cf [1], [3], [6], where this idea is directed to other applications.

It is quite surprising that it is so difficult to prove the intuitively obvious relations $f(m+1,n) > f(m,n) > f(m,n+1)$. There should be another formulation which makes this result obvious at a glance.

BIBLIOGRAPHY

1. Bellman, R., "An Introduction to the Theory of Dynamic Programming", RAND Report No. R-245, 1953.

2. ————————, "Dynamic Programming of Continuous Processes", RAND Report No. R-271, 1954.

3. ————————, "Computational Problems in the Theory of Dynamic Programming", Symposium on Numerical Analysis, August 1953, Santa Monica, California.

4. ————————, "Some Problems in the Theory of Dynamic Programming", Econometrica, January 1954, pp 37-48.

5. ————————, "Some Functional Equations in the Theory of Dynamic Programming", (to appear).

6. ————————, "Monotone Convergence in Dynamic Programming and the Calculus of Variations", Proc. Nat. Acad. Sci., (to appear).

7. ————————, "On the Equation of Optimal Inventory", (to appear).

8. Bellman, R., Harris, T. E., and Shapiro, H. N., "Some Functional Equations Occurring in the Theory of Decision Processes", RAND Paper No. P-382, 1953.

9. Johnson, S., and Karlin, S., "A Bayes Model in Sequential Design", RAND Paper No. P-328, 1954.

10. Karlin, S., "A Mathematical Treatment of Learning Models", RAND Research Memorandum 921, 1952.

11. Mahalanobis, P. C., "A Sample Survey of the Acreage Under Jute in Bengal", (with discussion of the planning of experiments), Sankhya, Volume 4, 1940, pp 511-531.

12. ————————————, "On Large-Scale Sample Surveys", Philos. Trans. Roy. Soc. London, Ser. B, Volume 231, 1944, pp 329-451.

13. Robbins, H., "Some Aspects of the Sequential Design of Experiments", Bul. Amer. Math. Soc., Volume 58, 1952, pp 527-536.

14. Thompson, W. R., "On the Likelihood That One Unknown Probability Exceeds Another in View of the Evidence of Two Samples", Biometrika, Volume 25, 1933, pp 285-294.

15. ————————, "On The Theory of Apportionment", Amer. Jour. Math., Volume 57, 1935.

md